Impact-Aware Manipulation by Dexterous Robot Control and Learning in Dynamic Semi-Structured Logistic Environments

Data Management Plan

Dissemination level	Public (PU)
Work package	WP6: Management
Deliverable number	D6.3
Version (D = Draft, F = Final)	F – 1.0
Submission date	29/06/2020
Due date	30/06/2020

www.i.am-project.eu

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 871899.

Authors

Authors in alphabetical order		
Name	**Organisation**	**Email**
Maarten JONGENEEL	TU/e	m.j.jongeneel@tue.nl
Jos DEN OUDEN	TU/e	j.h.v.d.ouden@tue.nl
Alessandro SACCON	TU/e	a.saccon@tue.nl

Control sheet

Version history			
Version	**Date**	**Modified by**	**Summary of changes**
0.1	13/05/2020	Jos DEN OUDEN	ToC & first contents
0.11	05/06/2020	Jos DEN OUDEN, Alessandro Saccon	New layout & updated all sections
0.12	26/06/2020	Jos DEN OUDEN	Ready for internal review
0.2	28/06/2020	Alessandro SACCON	Pre-final version ready for peer-review
0.5	29/06/2020	Maarten JONGENEEL	Pre-final version (revised)
0.7	29/06/2020	Abderrahmane KHEDDAR	Peer-review comments addressed
1.0	29/06/2020	Alessandro SACCON	Revised version ready for submission, quality check

Peer reviewers		
	Reviewer name	**Date**
Reviewer 1	Abderrahmane KHEDDAR	29/06/2020
Reviewer 2		

Legal disclaimer

TABLE OF CONTENTS

ABBREVIATIONS

Abbreviation	Definition
DMP	Data Management Plan
DOI	Digital Object Identifier
EC	European Commission
HDF	Hierarchical Data Format
PU	Public
SEO	Search Engine Optimization
WP	Work Package

EXECUTIVE SUMMARY

This deliverable D6.3 Data Management Plan, aims at providing an overview of the plan of the dataset generated in WP1 (I.Model), that is to be used by the entire consortium. This deliverable build upon the H2020 Online Manual on Data Management Plan, and deals with the topics of FAIR data, allocation of resources within the project dedicated to Data Management, data security, and ethical aspects. Since the I.AM. project mainly focuses within WP1 on the generation of the I.AM. dataset, we will describe how we intend to share it and the possible re-use within and outside the consortium. Further detailed technical implementation of the dataset will be described in D1.1 in M9 (September 2020), where we will explain the contents of the dataset, the location and the first description of how we ensure open access to general public.

This deliverable will be revised and updated at M25 (January 2022) in D6.5 "Data Management Plan (update D6.3)" and at M40 (April 2023) in D6.6 "Data Management Plan (second update D6.3)" and will receive updates from D1.1 as well.

1. INTRODUCTION

1.1. Project background

Europe is leading the market of torque-controlled robots. These robots can withstand physical interaction with the environment, including impacts, while providing accurate sensing and actuation capabilities. I.AM leverages this technology and strengthens European leadership by endowing robots to exploit intentional impacts for manipulation. I.AM focuses on impact aware manipulation in logistics, a new area of application for robotics which will grow exponentially in the coming years, due to socio-economical drivers such as booming of e-commerce and scarcity of labour.
I.AM relies on four scientific and technological research lines that will lead to breakthroughs in modelling, sensing, learning and control of fast impacts:

1. I.Model offers experimentally validated accurate impact models, embedded in a highly realistic simulator to predict post-impact robot states based on pre-impact conditions;
2. I.Learn provides advances in planning and learning for generating desired control parameters based on models of uncertainties inherent to impacts;
3. I.Sense develops an impact-aware sensing technology to robustly assess velocity, force, and robot contact state in proximity of impact times, allowing to distinguish between expected and unexpected events;
4. I.Control generates a framework that, in conjunction with the realistic models, advanced planning, and sensing components, allows for robust execution of dynamic manipulation tasks.

This integrated paradigm, I.AM, brings robots to an unprecedented level of manipulation abilities. By incorporating this new technology in existing robots, I.AM enables shorter cycle time (10%) for applications requiring dynamic manipulation in logistics. I.AM will speed up the take-up and deployment in this domain by validating its progress in three realistic scenarios: a bin-to-belt application demonstrating object tossing, a bin-to-bin application object fast boxing, and a case depalletizing scenario demonstrating object grabbing.

1.2. Data management

The I.AM. consortium agreed to join the H2020 Open Research Data Pilot. In line with such Pilot expectations, a first Data Management plan (D6.3) is produced by M6 mapping the main dataset generated by the project, originating from WP1. This dataset is also planned to be shared with research community and others, in M36 of the project. The Data Management Plan will be updated twice and revised before each validation phase to fine-tune it to the progresses of the research.

According to the I.AM. Grant Agreement article 29.3, the ORDP states that the consortium must comply with the following:

"Regarding the digital research data generated in the action ('data'), the beneficiaries must:

1. deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the following:
    (a) the data, including associated metadata, needed to validate the results presented in scientific publications, as soon as possible;
    (b) non applicable;
    (c) other data, including associated metadata, as specified and within the deadlines laid down in the 'data management plan' (see Annex 1);
2. provide information — via the repository — about tools and instruments at the disposal of the beneficiaries and necessary for validating the results (and — where possible — provide the tools and instruments themselves).

This does not change the obligation to protect results in Article 27, the confidentiality obligations in Article 36, the security obligations in Article 37 or the obligations to protect personal data in Article 39, all of which still apply.
As an exception, the beneficiaries do not have to ensure open access to specific parts of their research data under Point (a)(i) and (iii), if the achievement of the action's main objective (as described in Annex 1) would be jeopardised by making those specific parts of the research data openly accessible.
In this case, the data management plan must contain the reasons for not giving access."

1.3. Purpose of the deliverable

This deliverable D6.3 "Data Management Plan" provides an overview of the plan of the dataset generated in WP1, that is to be used by the entire consortium. We describe how we will share it and the possible re-use within and outside the consortium. The further detailed technical implementation of the dataset will be described in D1.1 in M9 (September 2020), where we will explain the contents of the dataset, the location, and the first description of how we ensure open access to general public. The contents of this document are based on H2020 DMP template.

This deliverable will be revised and updated at M25 (January 2022) in D6.5 ``Data Management Plan (update D6.3)'' and at M40 (April 2023) in D6.6 ``Data Management Plan (second update D6.3)'' and will receive updates from D1.1 as well.

1.4. Intended audience

The dissemination level of D6.3 is 'public' (PU) – meant for members of the Consortium (including Commission Services) and the general public. This document is intended to serve as an internal guideline for the entire I.AM. Consortium and provide the consortium's implementation plans regarding data management.

2. DATA MANAGEMENT PLAN

2.1. DATA SUMMARY

Although impact dynamics has been addressed in the robotics literature since its early days, there is no open dataset specifically addressing the recording and sharing of impact (and release) events in robotics.

I.AM. intends to create a dataset that will contain object-environment, robot-object, and robot-object-environment collisions data, according to the taxonomy illustrated in Figure 1. All datasets will be recorded in lab environments, providing a controlled environment.

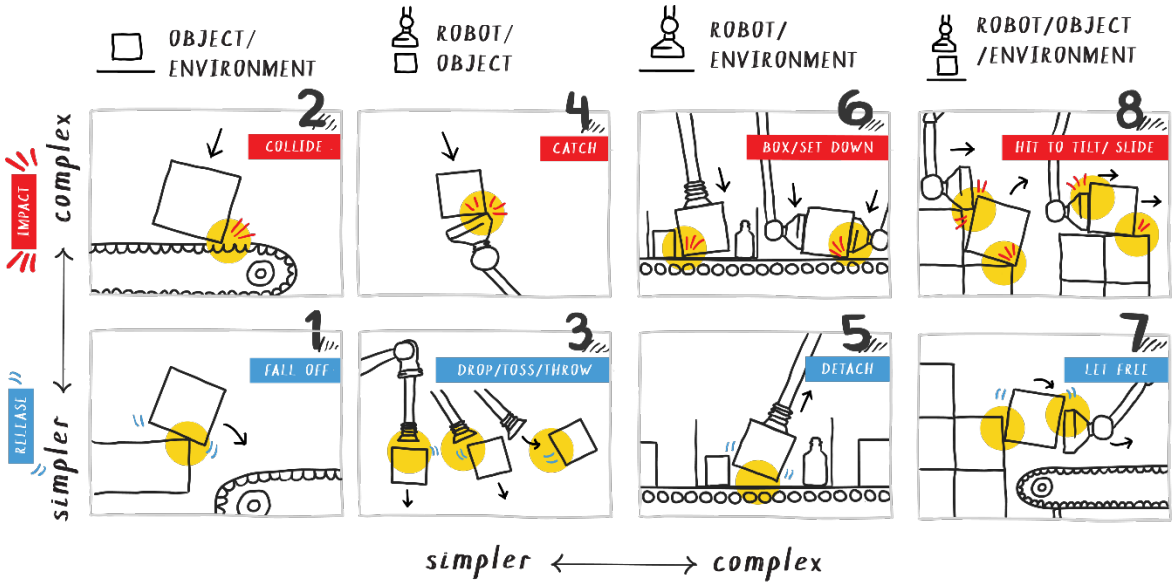


Figure 1:Preliminary illustration of I.AM. taxonomy. The complexity metric (levels 1 to 8) above will be used and expanded in deliverable D1.1 to create the I.AM. taxonomy, encompassing the description, nomenclature, and classification of dynamic contact events

The dataset that is being developed in WP1 is planned to be stored in HDF5 format. The first release of the dataset will be described in deliverable D1.1, which will go into more detail on the format of the data. An update of the dataset and findings after D1.1. (September 2020) will be again further reported in the updates of the Data Management Plan D6.5 (M25) and D6.6 (M40). Figure 2 shows an overview of the data and metadata that will be stored in the dataset. Typical recording is approximately 3 MB, without video footage and 70 MB with video footage. A typical dataset will contain around 300 measurements with a size of approximately 700MB – 1 GB each. The consortium plans to open the dataset with video footage, so it is more easily understandable how the tests were performed and expect to produce >10TB of datasets in this way.

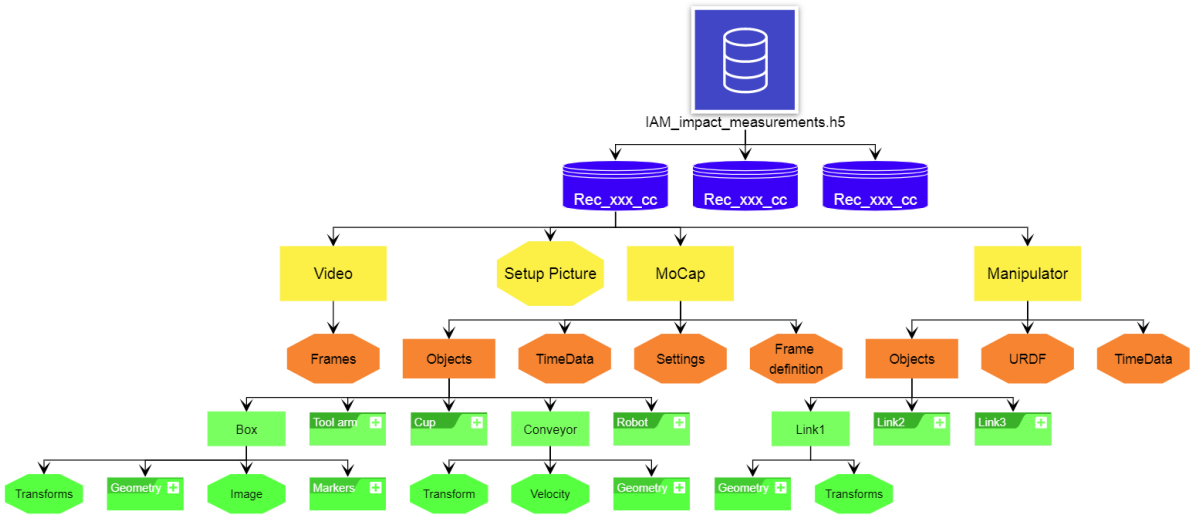


Figure 2: An example of dataset structure using the HDF5 hierarchy data file format, for recoding box-conveyor impact data relevant for the first validation scenario (TOSS)

Data from robot force/torque sensors, joint torque sensors, joint encoders, high speed camera, and motion capture data are planned to be added or adapted to the current set up.
Data will be stored in HDF5 [1] with internal containers containing standard or well established and open formats such as .csv and .bag (ROS) files.
Metadata such as author, measurement date/time, setup location, pictures of the setup, sensors used in the setup, geometry of the robots, conveyer belt and objects and CAD files, test settings (of camera, Opti track, robot parameters, URDF files etc.) will be stored together with the original sensor data. Figure 3 shows an example of implementation in HDF5 view [2].

Additionally, the I.AM. project will also open 3 software modules that are being developed in the project: I.Learn, I.Sense, and I.Control. Each of these will be described in a separate deliverable and made available to public using GitHub or GitLab repositories in M42 (June 2023) of the project.

2.2. FAIR DATA

2.2.1. Making data findable, including provisions for metadata

In order to make the dataset available and findable for the public, the I.AM. project aims at setting up a server for sharing the dataset, similar to what done in other dataset sharing initiatives (e.g., the KIT Whole-Body Human Motion Database [3]). These databases use keywords to make the data *findable* and the data is split up in small discrete parts, so that it is also *usable*.
Currently, the I.AM. project is already utilizing the HDF5 format for storing the data locally for the following reasons (see also [1]):

- Versatile data model that can represent very complex, heterogeneous data objects and a wide variety of metadata through an unlimited variety of datatypes;
- Ready for high speed raw data acquisition;
- Portable and extensible with no limits on file size, allowing applications to evolve in their use of HDF5;
- Self-describing, requiring no outside information for applications to interpret the structure and content of a file;
- Robust software ecosystem of open source tools and applications for managing, manipulating, viewing, and analysing data;
- Architecturally independent software library that runs on a wide range of computational platforms (from laptops to massively parallel systems) and programming languages (including C, C++, Fortran 90, and Java interfaces);
- Advanced performance features that allow for access time and storage space optimizations through customizable product packaging, compression, and encryption
- Long-term data archiving solution.

Using the open source available HDF5 to REST-based API, h5serv, the data from these HDF5 files can easily be transported to the server and the data as well as metadata made available to public. h5serv is a web service that implements a REST-based web service for HDF5 data stores [4].
Regarding open source access, there are multiple converters available, such as Python, MATLAB and C++ making this format easily accessible to everybody.
Metadata, such as described in Section 2.1 will be added to the test files.
Naming conventions are to be further defined. An initial suggestion is given here, which will be updated in M9 in public deliverable D1.1 "Publication of I.AM. dataset" and subsequentially be updated in M36 for the update of that deliverable, namely D1.4 "Publication of I.AM. dataset (update D1.1)".
Since HDF5 files can consists of multiple recordings, the main file naming convention needs to describe at least the physical objects and the motion/action that has been tested, for other users to be clear what the contents of the file is. We propose the following naming convention to be used:

I_AM_XYZ_xxxscenario_vy.z,

with example

I_AM_MCO_TOSSscenario_v1.0

meaning "I.AM. dataset with Manipulator, Conveyer & Object in the TOSS scenario version 1.0".

I.AM. data versioning will follow the Major.Minor numbering rule, similar to software versioning systems (e.g. v2.1). Major data revision will indicate a change in the formation and/or content of the dataset. Minor revisions will rather involve quality improvement over existing data items. Only minor edits are expected along the project implementation although major revisions are possible beyond the end of the I.AM. project.

The subsequent recordings, within these HDF5 files, will then be adhering a naming convention similar to the following one:

Rec_YYYY-MM-DD_hh-mm-ss_MRV

with example

Rec_2020-05-27_16-24-59_MRV

where MRV being a combination of the following types of data: M = motion capture data of Manipulator, Conveyer and/or Object, R = Robot data, V = video.

The type of measurement data can provide an indication about the quality and accuracy of the measurement and is therefore important metadata to be shared. An example is shown in Figure 3:

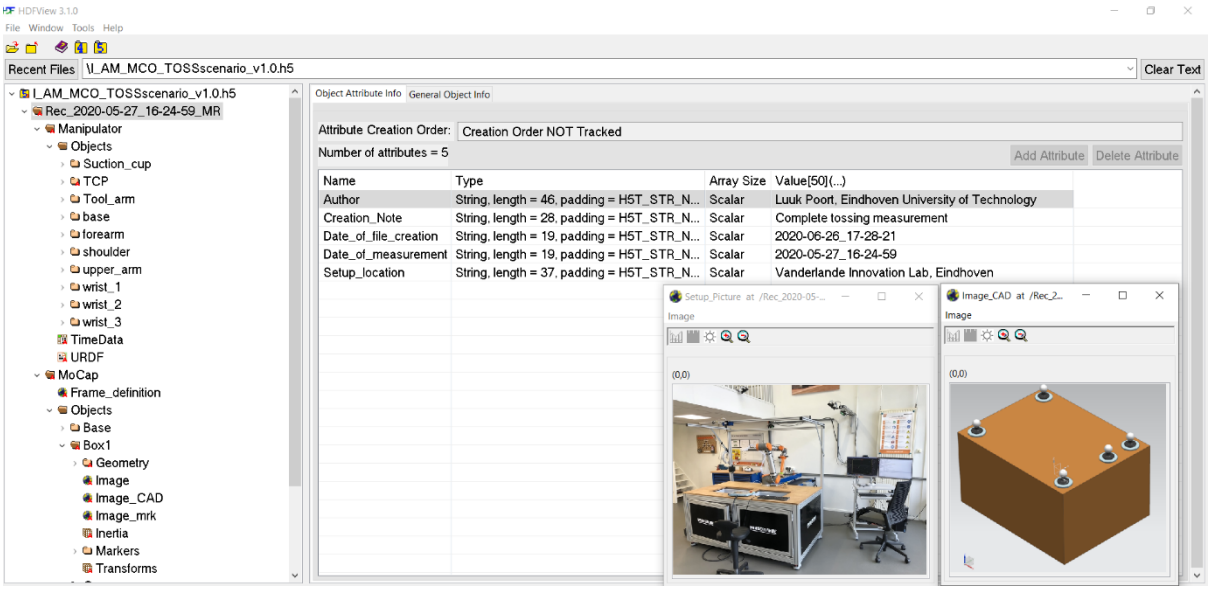


Figure 3: HDF5 view showing the metadata on the highest level (left). Other metadata (such as the setup picture, CAD images etc.) are captured in the lower branches of the structure and shown here (right) for reference as well.

Since each of the recordings will have a unique name and timestamp in the naming convention, this can be used to produce a Universally Unique Identifier (UUID).

The following generic keywords are planned to be utilized. This list will be updated throughout the project, when specific needs will require its revision:

- Authors and their affiliation

- Contributors and their affiliation
- Data location and persistent identifier
- Related/alternate identifiers
- Related journals, conferences, books and/or theses or other references
- Chosen license (for allowing the use of the specific dataset)

The following content keywords (also important for SEO) will be utilized:

- Impact
- Collision
- Impact-aware manipulation
- Impact-aware robotics
- Type of robot (Franka Emika Panda, KUKA iiwa, UR10…)
- Type of mocap (Optitrack…)
- Robot and impacts
- Robot and collisions
- Robots and dynamic manipulation
- Switching contact
- Impact and release
- Robotics motion control
- Robotics

In addition to this, complete project documentation will accompany the data to allow correct data interpretation and eventual experiment reproduction. This will include:

- Dataset overview – number of sub-datasets; status of documented data (complete or in progress); eventual plan of future update
- Methodological information – methods used for experimental design, data collection and data processing; instruments and software used; experimental conditions; quality assurance procedures performed on data
- Software and tools information – Name of tool/software; reference version; reference URL; eventual DOI.

2.2.2. Making data openly accessible

I.AM. aims to mainly provide the dataset that is to be produced in WP1 to the general public in a similar fashion as the KIT Whole-Body Human Motion Database [3]. Currently, I.AM. is also investigating the possibility to additionally provide the data to other platforms, such as Zenodo , since the search and metadata functionality of these open data platforms could improve the findability and accessibility to general public even further. Zenodo is a general-purpose open-access repository developed under the European OpenAIRE program and operated by CERN. It allows researchers to deposit data sets, research software, reports, and any other research related digital artefacts [5]. See also Annex 2 for more details on Zenodo.

Since the HDF5 format is supported by The HDF Group, a non-profit corporation whose mission is to ensure continued development of HDF5 technologies, and the continued accessibility of data stored in HDF. In keeping with this goal, the HDF libraries and associated tools are available under a liberal, BSD-like license for general use [6], [7]. Also, several relevant software tools (such as HDFview, Python/C++/MATLAB converters) are available. To guarantee accessibility, I.AM. will embed the versions of the software tools that are being used during the project and embed this with the dataset.

To store the impact dataset, the I.AM. partners have data repositories in place already at their institutes as well as there could be the option to store the data setting up a storage space such as done for I.AM. public website. Other certified repositories such as Zenodo are being considered. Zenodo, in REST API to easily upload data files and ensuring the relevant metadata for making data FAIR are provided [8]. Users can use their (existing) ORCID or GitHub accounts to get access to Zenodo, preventing the use of additional accounts and providing easy access (see Annex 2 for an overview of Zenodo). In addition, there are Python wrappers available for this API.
For the deliverable D1.1, a first repository will be in place for use within the consortium (and opened later to the public). Since there will be no personal or privacy sensitive data available in the WP1 dataset, the I.AM. consortium does not anticipate the use of a specific data access committee in order to provide access to the dataset.

According to the long-term dataset utility and potential limitations due to protection of personal data, different levels of confidentiality are considered within the I.AM. consortium:

- *Confidential to partner*. This option is applied when, regardless of the long-term value and scope for wider use, the dataset contains personal data that cannot be protected once disclosed. These include, among others, videos and images collected during the project tests.
- *Confidential to consortium* (including EC services). This option is applied for data containing confidential information (e.g. related to personal interviews) or those with no wide scope of use and long-term value.
- *Public*. This option is applied to most I.AM. datasets, particularly the WP1 dataset that is the main product of the project.

While datasets confidential to partner will be safely stored by the developing partner, all other datasets can be shared on the dedicated I.AM. repository.

Once the data is rich enough and all the tools for its exploitation are made operational efficiently, it will be submitted to be published and disseminated in the International Journal of Robotics Research, that welcomes such knowledge is a special paper section (Data paper).

2.2.3. Making data interoperable

The consortium will strive to collect and document the data in a standardized way to ensure that datasets can be correctly understood, interpreted, and re-used.

Currently, HDF5 format is considered as this is a standard that is widely used, and multiple converters are available to tools/languages, such as Python, MATLAB, C++.
A documentation describing the main variables included in the datasets will be provided in order to support the interpretation and re-use.
Standard vocabulary will be used for all data types present in the dataset to allow inter-disciplinary interoperability. In addition, the documentation will include a general glossary used to share information about the vocabulary and general methodologies employed for the generation of the dataset.

2.2.4. Increase data re-use (through clarifying licences)

Public data will be made available for re-use. To avoid any potential doubt, the consortium will attach specific licenses to the deposited data to define all conditions under which the work is provided under either open or restricted access.
In line with the project workplan, the impact motion database server will be made operational and the first WP1 dataset will be uploaded in M9 (September 2020). The database will be updated and finalized in M36 (December 2022) of the project.
Depending on the choice of data sharing, either by a dedicated repository set up by TU/e or by using another repository service, the I.AM. project will provide licensing (e.g., according to the Creative Commons). Although this will be further elaborated in D1.1, present I.AM. consortium's orientation-specific for the WP1 dataset is to licence the data as Creative Commons Attribution 4.0 International (CC BY 4.0)[a] allowing third parties to share and adapt data with no restrictions as long as attribution is provided. In case one partner would like to limit access to the uploaded data, alternative licenses will be selected among other CC licenses.

2.3. ALLOCATION OF RESOURCES

At this preliminary stage of the project, the only costs foreseen for data management are related to:

- the working time needed to set up and perform the data collection, including synchronisation of devices, and analysis activities
- the working time to setup local and shared data collection devices/servers
- the working time needed to write documentation, metadata, etc.
- the working time needed to set up and perform the data collection and analysis activities.

In WP1, task 1.2 is specifically set up for the generation of the I.AM. dataset with 16 PMs available for this. Main effort will be on the side of TU/e, with collaboration of partners EPFL, TUM, CNRS and Franka Emika.

[a] https://creativecommons.org/licenses/by/4.0/

The project coordinator and the project manager (TU/e) oversee the DM from both the scientific and technical perspective. TU/e will coordinate this effort (as part of WP1) and perform sample tests to maintain the quality of the dataset.

Validation and registration of datasets and metadata, as well as backing up data for sharing through open access repositories is the responsibility of the partner that generates the data in the WP. Each partner will identify a specific responsible person for each dataset. Quality control of this data is the responsibility of the relevant WP leader, supported by the Project Coordinator. Each partner should respect the policies set out in this DMP.

2.4. DATA SECURITY

As previously stated, each partner oversees backing up data that will be openly shared in the I.AM. repository.

We will consider publishing the dataset both on a dedicated I.AM. repository as well as to a public repository such as Zenodo, so to have an additional independent backup on two different locations/servers. This will guarantee long-term preservation in the unlikely event that one of the data repositories will cease operations. At the time of writing this deliverable, this decision is not final yet. More details will be provided in deliverable D1.1.

Generally, local backup system will be guaranteed during the project lifespan. Once uploaded on the I.AM. repository, data will also be stored in multiple online independent replicas. In case of the use of the I.AM. repository, TU/e also has multiple online replicas on its data servers available.

2.5. ETHICAL ASPECTS

Regarding the WP1 dataset, we do not expect any ethical or legal issues to arise.
Informed consent forms are separately made available to all partners through the Ethics deliverables D8.1 - H Requirements No.1 (M3) and D8.2 - POPD Requirements No.2 (M6).

3. CONCLUSION

This deliverable D6.3 provides a first data management plan for the I.AM. project in M6. It provides insight and considering the questions from the H2020 Online Manual on what, how and where to store open datasets that are required to be shared with general public.
The project will continue till M48 (December 2023) and this deliverable will be updated twice in M25 and M40.
Aside from this data management plan, there are also two other deliverables (D1.1. and D1.4) that also cover large part of how I.AM. will provide an open dataset. In those deliverables, we will expand more on the contents of the dataset and the motivation for this, whereas the data management plan deliverables will focus mainly on how to provide these datasets to the general public in line with the expected regulations and tools provides by open research platforms.

4. REFERENCES

[1] The HDF Group, "The HDF5® Library & File Format - The HDF Group." https://www.hdfgroup.org/solutions/hdf5/ (accessed Jun. 26, 2020).

[2] T. H. Group, "HDF® View - The HDF Group." https://www.hdfgroup.org/downloads/hdfview/ (accessed Jun. 26, 2020).

[3] "KIT Whole-Body Human Motion Database." https://motion-database.humanoids.kit.edu/list/motions/ (accessed Jun. 26, 2020).

[4] The HDF Group, "HDFGroup/h5serv: Reference service implementation of the HDF5 REST API." https://github.com/HDFGroup/h5serv (accessed Jun. 26, 2020).

[5] "Zenodo - about." https://about.zenodo.org/ (accessed Jun. 27, 2020).

[6] "HDF Group Licenses and Copyrights." https://support.hdfgroup.org/products/licenses.html (accessed Jun. 29, 2020).

[7] "Hierarchical Data Format - Wikipedia." https://en.wikipedia.org/wiki/Hierarchical_Data_Format (accessed Jun. 29, 2020).

[8] "Developers | Zenodo." https://developers.zenodo.org/ (accessed Jun. 27, 2020).

[9] "Zenodo - infrastructure." https://about.zenodo.org/infrastructure/ (accessed Jun. 27, 2020).

ANNEX 1: DATA MANAGEMENT PLAN TEMPLATE

The questions below are extracted from H2020 Online Handbook on Data Management Plan and used to set up this DMP.
This list is provided as reference to all the partners in I.AM. to provide insight on sharing data and publication and the process of FAIR data that should be linked to it.

DATA SUMMARY

- What is the purpose of the data collection/generation and its relation to the objectives of the project?
- What types and formats of data will the project generate/collect?
- Will you re-use any existing data and how?
- What is the origin of the data?
- What is the expected size of the data?
- To whom might it be useful ('data utility')?

FAIR DATA

Making data findable, including provisions for metadata

- Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?
- What naming conventions do you follow?
- Will search keywords be provided that optimize possibilities for re-use?
- Do you provide clear version numbers?
- What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

Making data openly accessible

- Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.
  Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.
- How will the data be made accessible (e.g. by deposition in a repository)?
- What methods or software tools are needed to access the data?
- Is documentation about the software needed to access the data included?
- Is it possible to include the relevant software (e.g. in open source code)?
- Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.
- Have you explored appropriate arrangements with the identified repository?
- If there are restrictions on use, how will access be provided?

- Is there a need for a data access committee?
- Are there well described conditions for access (i.e. a machine-readable license)?
- How will the identity of the person accessing the data be ascertained?

Making data interoperable

- Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?
- What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?
- Will you be using standard vocabularies for all data types present in your dataset, to allow inter-disciplinary interoperability?
- In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

Increase data re-use (through clarifying licences)

- How will the data be licensed to permit the widest re-use possible?
- When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.
- Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.
- How long is it intended that the data remains re-usable?
- Are data quality assurance processes described?

ALLOCATION OF RESOURCES

- What are the costs for making data FAIR in your project?
- How will these be covered? Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions).
- Who will be responsible for data management in your project?
- Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?

DATA SECURITY

- What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?
- Is the data safely stored in certified repositories for long term preservation and curation?

ETHICAL ASPECTS

- Are there any ethical or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).
- Is informed consent for data sharing and long-term preservation included in questionnaires dealing with personal data?

OTHER ISSUES

- Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?

FURTHER SUPPORT IN DEVELOPING YOUR DMP

The Research Data Alliance provides a Metadata Standards Directory[b] that can be searched for discipline-specific standards and associated tools.

The EUDAT B2SHARE[c] tool includes a built-in license wizard that facilitates the selection of an adequate license for research data.

Useful listings of repositories include:

Registry of Research Data Repositories[d]:
Some repositories like Zenodo[e], an OpenAIRE[f] and CERN collaboration), allow researchers to deposit both publications and data, while providing tools to link them.
Other useful tools include DMP online[g] and platforms for making individual scientific observations available such as ScienceMatters[h].

[b] http://rd-alliance.github.io/metadata-directory/
[c] https://b2share.eudat.eu/
[d] https://www.re3data.org/
[e] https://zenodo.org/
[f] https://www.openaire.eu/
[g] https://dmponline.dcc.ac.uk/
[h] https://www.sciencematters.io/

ANNEX 2: ZENODO PLATFORM

What is Zenodo?

Built and developed by researchers, to ensure that everyone can join in Open Science.

The OpenAIRE project, in the vanguard of the open access and open data movements in Europe was commissioned by the EC to support their nascent Open Data policy by providing a catch-all repository for EC funded research. CERN, an OpenAIRE partner and pioneer in open source, open access and open data, provided this capability and Zenodo was launched in May 2013.

In support of its research programme CERN has developed tools for Big Data management and extended Digital Library capabilities for Open Data. Through Zenodo these Big Science tools could be effectively shared with the long tail of research.

Why use Zenodo?

- Safe — your research is stored safely for the future in CERN's Data Centre for as long as CERN exists.
- Trusted — built and operated by CERN and OpenAIRE to ensure that everyone can join in Open Science.
- Citeable — every upload is assigned a Digital Object Identifier (DOI), to make them citable and trackable.
- No waiting time — Uploads are made available online as soon as you hit publish, and your DOI is registered within seconds.
- Open or closed — Share e.g. anonymized clinical trial data with only medical professionals via our restricted access mode.
- Versioning — Easily update your dataset with our versioning feature.
- GitHub integration — Easily preserve your GitHub repository in Zenodo.
- Usage statistics — All uploads display standards compliant usage statistics
- You can use either a dedicated Zenodo account, or existing GitHub or ORCID accounts to get access to the Zenodo platform.

You can login using either a dedicated Zenodo account or using your (existing) GitHub and/or ORCID account (see Figure 4).

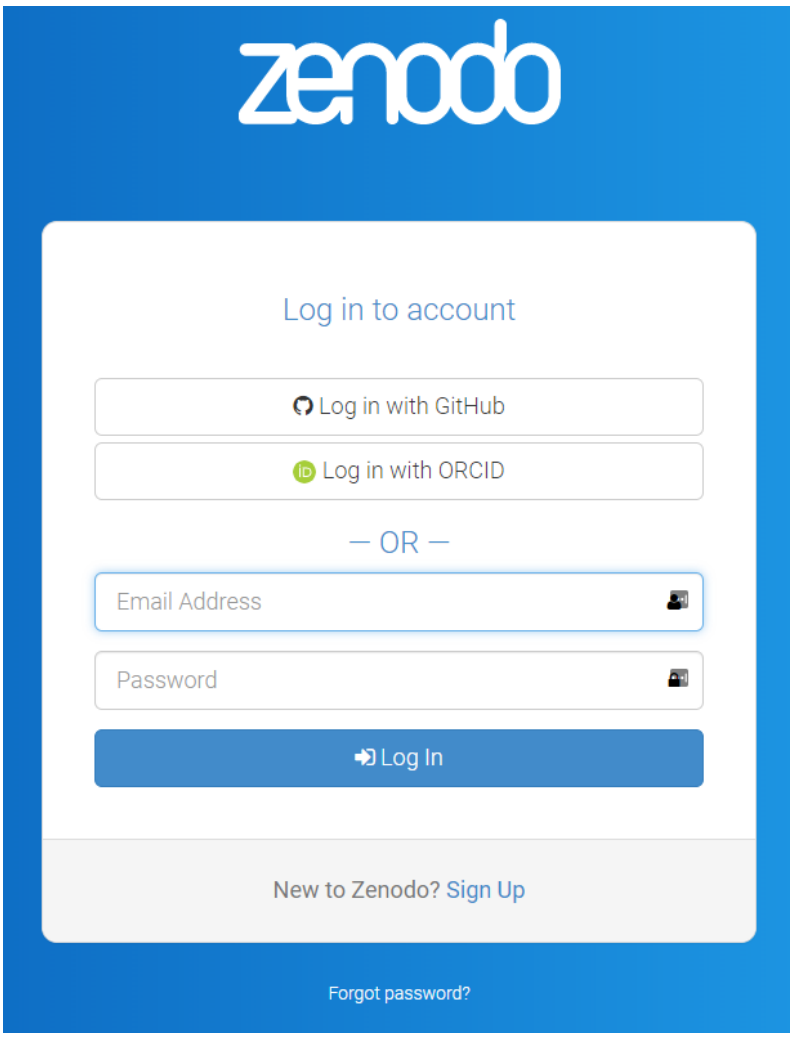


Figure 4: Login screen of Zenodo platform

Upload files (publications, datasets, software etc.)

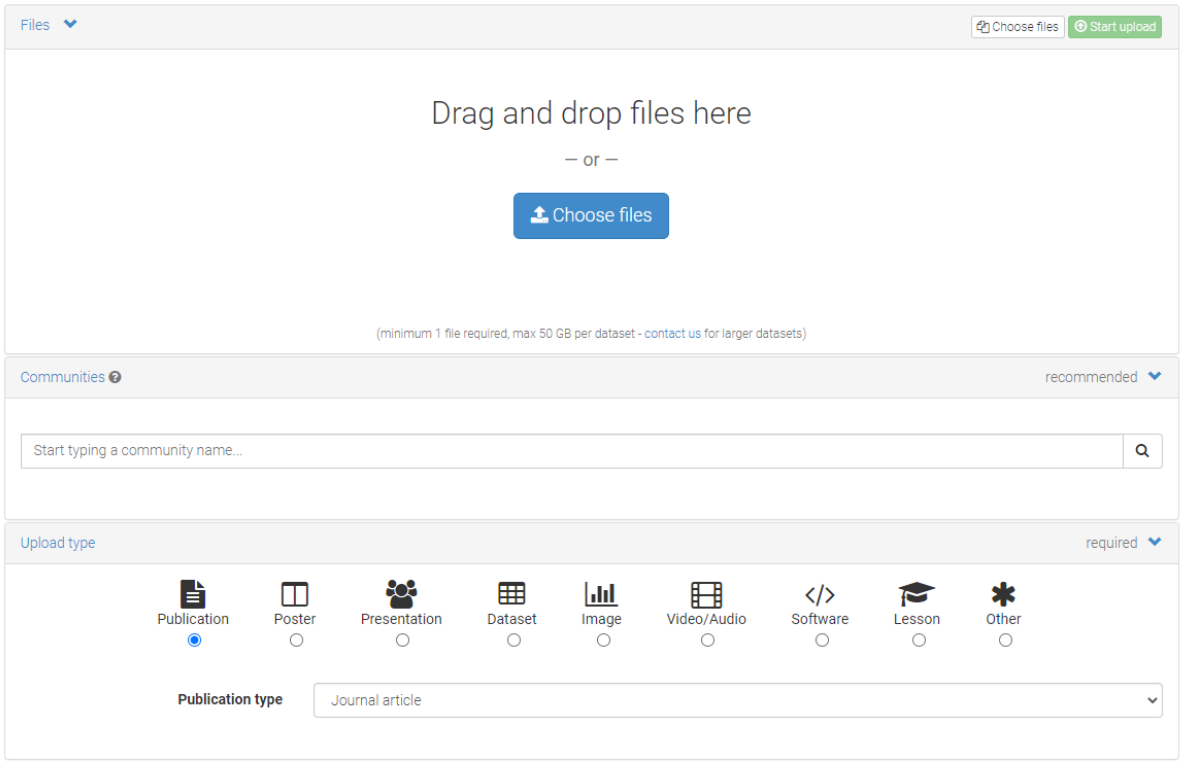
The figure below (Figure 5) show the requested steps and information in order to upload your publications, datasets, software etc. to Zenodo and make it immediately in line with the FAIR data framework.

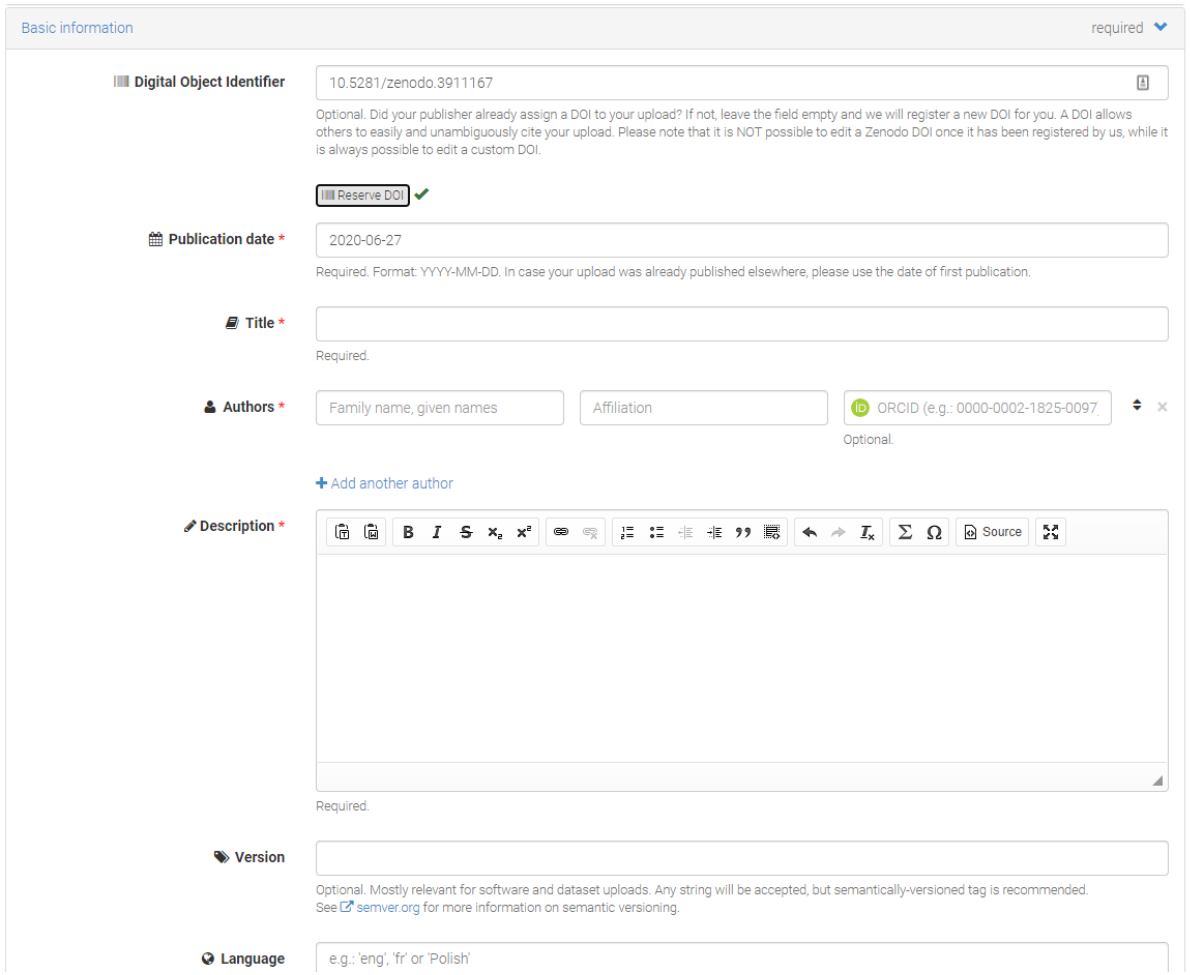
When uploading data, the I.AM. partners should be asked to choose among four main options:

- Open Access. This is the highly recommended option which provides free access and rights to data
- Embargoed Access. This option will be applied in case of data underpinning publications. Data will indeed be deposited as soon as possible but open access will be provided only once such data have been published in a scientific paper to preserve the authorship of all authors involved. In such case, information about data will be published and details of when the data will become available will be included in the metadata.
- Restricted Access. This option, although not recommended, will be adopted for those data whose access to should be monitored and approved by the depositor if certain requirements to be defined are met.
- Closed Access. This option will be adopted for all datasets whose confidentiality is limited to consortium and EC services.

Although the embargoed or closed access option provided by Zenodo could be a valid option, the consortium agrees that research data linked to exploitable results will not be deposited to avoid compromising their protection or commercialisation prospects. As clearly specified on Zenodo security provisions, "closed access is not suitable for secret or confidential data" since these are "stored unencrypted and may be viewed by Zenodo operational staff" [9].
Visibility and access to publicly shared datasets will be facilitated by Zenodo metadata and search facility as well as to the automatic link to both OpenAIRE and project Cordis project page.

Automatically upload large amount of data (using the API):
Note also that there is a public API available, which you can use in order to upload large amounts of data automated, instead of 1 by 1 manually (see [8]).


New upload
Instructions: (i) Upload minimum one file or fill-in required fields (marked with a red star). (ii) Press "Save" to save your upload for editing later. (iii) When ready, press "Publish" to finalize and make your upload public.
Files
Choose files
Start upload
Drag and drop files here
– or –
Choose files
(minimum 1 file required, max 50 GB per dataset - contact us for larger datasets)
Communities
recommended
Start typing a community name...
Upload type
required
Publication
Poster
Presentation
Dataset
Image
Video/Audio
Software
Lesson
Other
Publication type
Journal article


Basic information
required
Digital Object Identifier
10.5281/zenodo.3911167
Optional. Did your publisher already assign a DOI to your upload? If not, leave the field empty and we will register a new DOI for you. A DOI allows others to easily and unambiguously cite your upload. Please note that it is NOT possible to edit a Zenodo DOI once it has been registered by us, while it is always possible to edit a custom DOI.
Reserve DOI
Publication date *
2020-06-27
Required. Format: YYYY-MM-DD. In case your upload was already published elsewhere, please use the date of first publication.
Title *
Required.
Authors *
Family name, given names
Affiliation
ORCID (e.g.: 0000-0002-1825-0097
Optional.
Add another author
Description *
Source
Required.
Version
Optional. Mostly relevant for software and dataset uploads. Any string will be accepted, but semantically-versioned tag is recommended.
See semver.org for more information on semantic versioning.
Language
e.g.: 'eng', 'fr' or 'Polish'

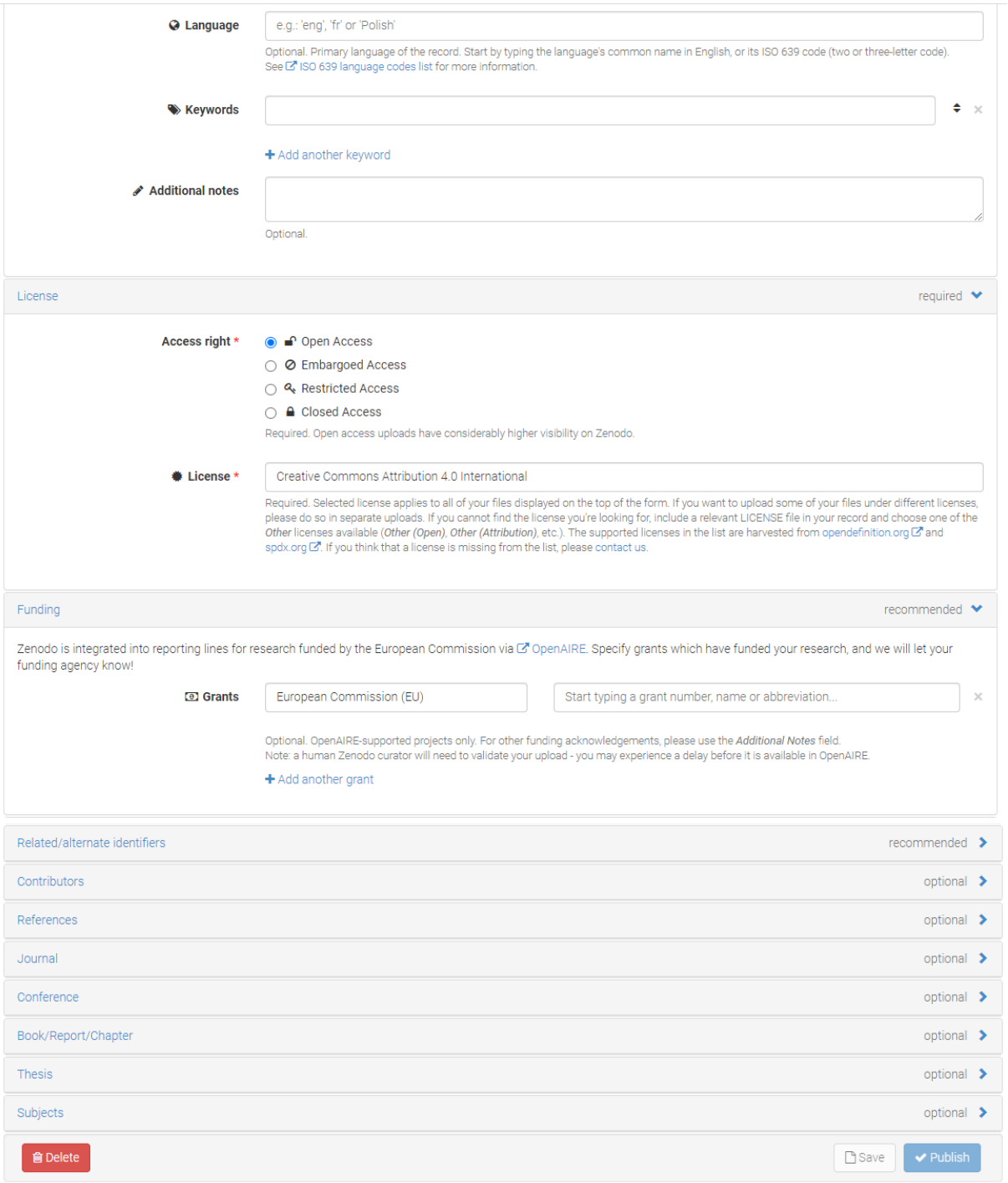


Figure 5: Zenodo upload screen with metadata